Optimal change-point estimation in inverse problems

Michael H. Neumann
Weierstrass Institute for Applied Analysis and Stochastics, Berlin

Abstract. We develop a method of estimating change–points of a function in the case of indirect noisy observations. As two paradigmatic problems we consider deconvolution and errors-in-variables regression. We estimate the scalar products of our indirectly observed function with appropriate test functions, which are shifted over the interval of interest. An estimator of the change point is obtained by the extremal point of this quantity. We derive rates of convergence for this estimator. They depend on the degree of ill-posedness of the problem, which derives from the smoothness of the error density. Analyzing the Hellinger modulus of continuity of the problem we show that these rates are minimax.

1991 Mathematics Subject Classification. Primary 62G05; secondary 62G20
Key words and phrases. Change–point estimation, inverse problems, indirect observations, deconvolution, errors-in-variables regression, optimal rates of convergence

1. Introduction

Change-point estimation has often been studied in the regression context. There are many practical motivations why one is interested in knowing such points of rapid change. Sometimes there is real scientific interest in these point, but one can also exploit knowledge about them for estimation purposes themselves.

The simplest case is that of a single jump of an otherwise smooth function. The optimal rate at which then a change-point can be estimated is known to be n^{-1}. Korostelev (1987) derives an optimal method in the Gaussian white noise model, which can also be applied in the usual nonparametric regression setting. Another popular approach is based on the analysis of differences of certain kernel estimators; see, e.g., Yin (1988), Müller (1992), Hall and Titterington (1992) and Wu and Chu (1993). Wang (1995) considers a closely related method based on wavelets. It can be shown that one can achieve the optimal rate of convergence also by the kernel-based method, provided one uses an appropriate, necessarily discontinuous kernel.

In the present paper we study this problem in the context of ill-posed inverse problems. Such problems arise when we can observe an object of interest only indirectly. Typical settings are deconvolution, errors-in-variables regression, estimation of mixing densities, image blur models and image reconstruction in computerized tomography. The quality at which a function can be estimated from such indirect, noisy observations depends on the degree of ill-posedness of the problem. For example, deconvolution becomes harder as smoothness of the error distribution increases. Most of the available results focus on the estimation of functions with homogeneous smoothness; see Hall (1990), Zhang (1990), Fan (1991) and Fan and Truong (1993). However, in practical applications one is often confronted with functions that have quite inhomogeneous smoothness characteristics: they are quite smooth on one part of the domain, but much less regular on another part. In such situations usual linear smoothing methods, which apply a global degree of smoothing, are no longer appropriate. Locally adaptive methods for estimating a function in the setting of indirect observations are developed by Donoho (1992) on the basis of Wavelet-Vaguelette Decompositions. However, the author is not aware of any work on change-point estimation in this context.

Here we focus on change-point estimation in one-dimensional inverse problems, as opposed to the more complex problem of edge estimation in higher dimensions. We study a quite general method and show its applicability for two paradigmatic problems, deconvolution and errors-in-variables regression.

Our motivation is at least twofold: First, knowledge of the location of a jump is helpful when one intends to estimate the function itself. One can then use this information and apply one-sided estimation techniques around this point. Second, one might be interested in estimating the support of a density, which amounts to the estimation of change-points in the case of a sharp boundary. Finally, there exist many interesting higher-dimensional inverse problems like image deblurring or density estimation from computer tomography data. We hope that the methodology developed in this article can be carried over to edge detection in these important problems.

We propose a method which is similar to well-established kernel methods for direct observations. Starting from an appropriate compactly supported function φ we define test functions $\varphi_\vartheta(.) = \varphi((.-\vartheta)/h_n)$, where ϑ varies in the interval of interest and the bandwidth h_n tends to zero at a certain rate. Our method is based on the idea that the scalar products of φ_ϑ and a discontinuous function attain their maximum value in modulus at some ϑ close to the change-point ϑ_0. We define empirical versions of these scalar products and take that value of ϑ, which maximizes them in absolute value, as an estimator of ϑ_0. We show that this estimator converges with the rate $n^{-1/(\alpha+3/2)}$ to ϑ_0, where α is the degree of ill-posedness of the inverse problem. Analyzing the Hellinger modulus of continuity of the problem we show that this is actually the optimal rate of convergence.

2. Change-point estimation in the deconvolution problem

Suppose we have n i.i.d. random variables $X_1, \cdots, X_n$ distributed according to a density f_X. However, we do not observe the X_i's directly, but

$$Y_i = X_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where the ε_i's are i.i.d. with density f_ε, also independent of the X_i's.

There already exists a considerable body of literature on estimating the density f_X in model (2.1); e.g. Carroll and Hall (1988) and Fan (1991). The first of these papers also describes some practical applications. It turns out that f_X can be more or less successfully estimated on the basis of observations $\{Y_i\}$, where the rate at which an optimal estimator converges to f_X depends on the smoothness of both f_X and f_ε. Here, a smooth f_X and a rough f_ε are most favorable.

In ordinary change-point estimation problems, i.e. in the presence of direct observations, a widely used method is to take just that point, where the difference of two-sided kernel estimators attains its maximum in absolute value; cf. Yin (1988), Müller (1992) , Wu and Chu (1993) and Wang (1995). (Strictly speaking, Wang uses wavelets rather than the difference of two kernels; however he did not actually exploit their special properties like orthonormality, which distinguish them from the kernel difference approach.)

For the sake of a clear presentation of the main idea we consider the simplest case of a single jump, whose height is bounded away from zero. It will become clear how this method can also be used for the detection of an unknown number of jumps, which are separated from each other by at least an arbitrarily small, but fixed constant.

We will assume that the density f_X is a member of the class $\mathcal{F}$, where, for $a < b$,

$$\begin{aligned} \mathcal{F} \;=\; \{f| \quad & |f(x)| \le C_1 \quad \forall x, \quad \text{and for some } \vartheta_0 \in [a,b] : \\ & |f(\vartheta_0-) - f(\vartheta_0+)| \ge C_2 \quad \text{and} \quad |f(x) - f(y)| \le C_3|x-y|, \quad \text{if } \vartheta_0 \notin [x,y]\}. \end{aligned}$$

Here and in the following C, C_1, C_2, ... denote generic constants. Our basic assumption on the error density f_ε is, for some $0 < C_4 \le C_5 < \infty$ and $\alpha \ge 1$, that

(A1) $\quad C_4(1 + |\omega|)^{-\alpha} \le |\widehat{f}_\varepsilon(\omega)| \le C_5(1 + |\omega|)^{-\alpha} \quad \forall \omega \in \mathbb{R},$

where $\widehat{g}(\omega) = \int g(x)\exp(ix\omega)\,dx$ denotes the Fourier transform of a function $g \in L_1(\mathbb{R})$. This assumption basically means that that f_ε has about α derivatives. It turns out that (A1) with an appropriate value of α is satisfied for gamma distributions with shape parameter α, which contain for $\alpha = 1$ the exponential and for $\alpha = \vartheta/2$ the chi-square distribution with ϑ degrees of freedom as special cases. Another example, which satisfies (A1) with $\alpha = 2$, is the Laplace (double exponential) distribution.

The basic idea of our method is as follows. Assume for definiteness that $f_X(\vartheta_0-) \geq f_X(\vartheta_0+) + C_2$. Then f_X can be written as

$$f_X \;=\; f \,+\, h, \tag{2.2}$$

where $h(\vartheta_0-) \geq h(\vartheta_0+) + C_2$, $h(x) \geq 0$ for $x < \vartheta_0$, $h(x) \leq 0$ for $x > \vartheta_0$, and f and h are Lipschitz continuous on $\mathbb{R}$ and $\mathbb{R} \setminus \{\vartheta_0\}$, respectively. Let φ be some nonzero function with $\varphi(x) \geq 0$, if $x < 0$, and $\varphi(x) \leq 0$, if $x > 0$. The functions $\varphi_\vartheta(.) = \varphi((. - \vartheta)/h_n)$ with $\vartheta \in [a,b]$ and $h_n \to 0$ can be used for "scanning" f_X for a discontinuity. In particular, (φ_ϑ, h) will attain its maximum at $\vartheta = \vartheta_0$, whereas (φ_ϑ, f) is of smaller order because of the Lipschitz continuity of f and $h_n \to 0$. Our method amounts to estimating the scalar products (φ_ϑ, f_X) . An ideal choice of φ is such, that the ratio of the contrast $(\varphi_{\vartheta_0}, f_X) - (\varphi_\vartheta, f_X)$ and the noise $\sqrt{var(\widehat{(\varphi_{\vartheta_0}, f_X)} - \widehat{(\varphi_\vartheta, f_X)})}$, where $\widehat{(\varphi_\vartheta, f_X)}$ is an appropriate estimate of (φ_ϑ, f_X), is maximized. In ordinary change-point estimation, i.e. with *direct* noisy observations, this goal is achieved by a function φ with a discontinuity at zero. However, the situation changes in the context of *indirect* observations. To embed the deconvolution problem in the general frame of ill-posed inverse problems, define the bounded linear operator $K: \; L_1(\mathbb{R}) \to L_1(\mathbb{R})$ with

$$(Kg)(x) \;=\; \int g(x-y) f_\varepsilon(y)\, dy.$$

Deconvolution is ill-posed, since K does not have a bounded inverse. Since we are only given observations in the image of K, estimation of (φ_ϑ, f_X) amounts to estimating $(\gamma_\vartheta, K f_X)$ for some function γ_ϑ. Now it turns out that a discontinuous function φ_ϑ leads to a function γ_ϑ with an unbounded L_2-norm. Since this makes the following statistical analysis impossible, we have to look for better alternatives.

For the function φ we will require that

(A2) (i) φ has compact support,
(ii) φ is an odd function with $\varphi(x) \geq 0$, if $x < 0$,
(iii) $\|\widehat{\varphi}(\omega)(1+|\omega|)^{\alpha+2}\|_i < \infty$ for $i = 1, 2,$
(iv) $\varphi'(0) \neq 0.$

Note that (iv) of (A2) is essential. It takes the role of the discontinuity of φ in the case of direct observations, and provides the optimal ratio between contrast and noise in our case.

Let $\varphi_\vartheta(x) = \varphi((x-\vartheta)/h_n)$. The sequence of equalities

$$\begin{aligned}(\varphi_\vartheta, g) \;&=\; (2\pi)^{-1} \int \mathcal{R}e(\widehat{\varphi}_\vartheta(\omega)\widehat{g}(\omega))\, d\omega \;=\; (2\pi)^{-1} \int \mathcal{R}e(\widehat{\varphi}_\vartheta(\omega)/\widehat{f}_\varepsilon(\omega) \;\; \widehat{g}(\omega)\widehat{f}_\varepsilon(\omega))\, d\omega \\ &=\; \mathcal{R}e\left(\mathcal{F}^{-1}(\widehat{\varphi}_\vartheta/\widehat{f}_\varepsilon), g * f_\varepsilon\right) \;=\; (\mathcal{R}e(\mathcal{F}^{-1}(\widehat{\varphi}_\vartheta/\widehat{f}_\varepsilon)), g * f_\varepsilon), \end{aligned} \tag{2.3}$$

which is valid for all $g \in L_1(\mathbb{R})$, motivates us to define

$$\gamma_\vartheta \;=\; \mathcal{R}e\left(\mathcal{F}^{-1}(\widehat{\varphi}_\vartheta/\widehat{f}_\varepsilon)\right), \tag{2.4}$$

where $\mathcal{F}^{-1}(g)(x) = (2\pi)^{-1} \int g(\omega)\exp(-i\omega x)\, d\omega$ denotes the inverse Fourier transform.
We consider

$$\widetilde{\alpha}_\vartheta \;=\; n^{-1} \sum_{i=1}^{n} \gamma_\vartheta(Y_i), \tag{2.5}$$

which is an unbiased estimator of (φ_ϑ, f_X). Let

$$\alpha_\vartheta \;=\; E\widetilde{\alpha}_\vartheta \;=\; (\varphi_\vartheta, f_X). \tag{2.6}$$

A first insight into the essential properties of the $\widetilde{\alpha}_\vartheta$'s is given by the following lemma.

Lemma 2.1. *Assume (A1), (A2) and* $h_n \to 0$. *Then, for n large enough,*

$$(i) \qquad |\alpha_{\vartheta_0}| \;-\; |\alpha_\vartheta| \;\geq\; \begin{cases} C_6|\vartheta_0 - \vartheta|^2/h_n \;-\; C_7 h_n|\vartheta_0 - \vartheta|, & \text{if } |\vartheta_0 - \vartheta| \leq h_n, \\ C_8 h_n, & \text{if } |\vartheta_0 - \vartheta| > h_n \end{cases},$$

$$(ii) \qquad \sqrt{var(\widetilde{\alpha}_{\vartheta_0} \;-\; \widetilde{\alpha}_\vartheta)} \;=\; O\left(n^{-1/2}|\vartheta_0 - \vartheta| h_n^{-\alpha-1/2}\right), \quad \text{if } |\vartheta_0 - \vartheta| \leq h_n$$

$$(iii) \qquad \sqrt{var(\widetilde{\alpha}_\vartheta)} \;=\; O\left(n^{-1/2} h_n^{-\alpha+1/2}\right)$$

hold uniformly in $\vartheta \in [a, b]$ *and* $f_X \in \mathcal{F}$.

First, observe that, according to (i) and (iii) of the above lemma, the noise is of smaller order of magnitude than the contrast in the case of $|\vartheta - \vartheta_0| > h_n$, if $h_n \gg n^{-1/(2\alpha+1)}$. This will imply that such values of ϑ can be excluded with very high probability as values for $\widehat{\vartheta}_0$. Further we infer from (i), that some positive contrast is guaranteed, if $|\vartheta_0 - \vartheta| > (C_6/C_5)h_n^2$. Furthermore, the noise increases linearly in $|\vartheta_0 - \vartheta|$, whereas the contrast grows quadratically. Both quantities are equated (in order) at $|\vartheta_0 - \vartheta| \asymp n^{-1/2} h_n^{-\alpha+1/2}$. This already gives some hint that ϑ_0 can be estimated with the rate $h_n^2 + n^{-1/2} h_n^{-\alpha+1/2}$, which is optimized for $h_n \asymp n^{-1/(2\alpha+3)}$.
Let

$$\widehat{\vartheta}_0 \in \arg \max_{\vartheta \in [a,b]} \{|\widetilde{\alpha}_\vartheta|\}. \tag{2.7}$$

The following theorem states that $\Delta_n = n^{-1/(\alpha+3/2)}$ is actually the rate of convergence for this estimator.

Theorem 2.1. *Assume (A1), (A2) and* $h_n \asymp n^{-1/(2\alpha+3)}$. *Then*

$$\sup_{f_X \in \mathcal{F}} \left\{ E_{f_X} \left(\Delta_n^{-1}(\widehat{\vartheta}_0 \;-\; \vartheta_0) \right)^2 \right\} \;\leq\; C.$$

Remark 1.
1) As can be seen in the proof of this theorem, we have to choose the bandwidth h_n exactly of this order. If we increase h_n, then we blur the information about ϑ_0 too much. On the other hand, recognition of a signal becomes more difficult as the amount of localization increases. Hence, if we choose h_n smaller we would increase the magnitude of noise in $\widetilde{\alpha}_\vartheta$, which would also lead to a worse rate.

However, in sharp contrast to bandwidth selection in curve estimation, the optimal rate for h_n does not depend on any unknown smoothness parameters of the function to be investigated. It depends only on the degree of ill-posedness α of the problem, which we know exactly since we know f_ε.

2) For simplicity of presentation we have only considered the simplest case of a single discontinuity. Proceeding as in Wu and Chu (1993) one can easily extend this method to the case of multiple jumps, also to the case of an unknown number of them.

3. Change-point estimation in errors-in-variables regression

Nonparametric regression with errors in the independent variables forms another instance of an ill-posed inverse problem. It is well-known that then the estimation of the regression function is harder than in the case of exactly known regressors. This fact is underlined by results in Fan and Truong (1993), who derived minimax rates in errors-in-variables regression.

Let (X, Z) be a pair of random variables and let $m(x) = E(Z \mid X = x)$. However, we do not observe X directly, but $Y = X + \varepsilon$, where ε is some stochastic measurement error. Suppose, we are given a sample of n observations (Y_i, Z_i), we arrive at the following basic model:

$$Z_i \;=\; m(X_i) \;+\; \eta_i, \tag{3.1}$$

$$Y_i \;=\; X_i \;+\; \varepsilon_i. \tag{3.2}$$

Let f_X denote the density of X. To make the estimation problem identifiable, we assume to know the density f_ε of the i.i.d. errors ε_i, which are independent of the (X_i, Z_i)'s.

Now we assume that the function m has a single jump at $\vartheta_0 \in [a, b]$, and is smooth otherwise. Again we intend to estimate ϑ_0 as accurate as possible.

As a functional, which is aimed at drawing the essential information about ϑ_0 from the sample $\{(Y_i, Z_i)\}$ we consider the quantity

$$\widetilde{\beta}_\vartheta \;=\; n^{-1} \sum_{i=1}^{n} \gamma_\vartheta(Y_i) Z_i, \tag{3.3}$$

where γ_ϑ was defined by (2.4).

Now we have

$$\begin{aligned}
\beta_\vartheta &= E\widetilde{\beta}_\vartheta \\
&= E\gamma_\vartheta(X+\varepsilon)m(X) \\
&= \int\int \gamma_\vartheta(x+\varepsilon)m(x)f_X(x)f_\varepsilon(\varepsilon)\,dx\,d\varepsilon \\
&= \int \gamma_\vartheta(x) \int m(x-\varepsilon)f_X(x-\varepsilon)f_\varepsilon(\varepsilon)\,d\varepsilon\;dx \\
&= (\gamma_\vartheta, (mf_X) * f_\varepsilon) \\
&= (\varphi_\vartheta, mf_X).
\end{aligned} \tag{3.4}$$

Hence, $\widetilde{\beta}_\vartheta$ primarily aims at detecting a jump in $m(x)f_X(x)$ at ϑ. However, it can also serve for the detection of change points in m, if

(A3) f_X is uniformly Lipschitz, $f_X(x) \geq C$ for $x \in [a-\delta, b+\delta]$ and some $\delta > 0$.

Now, $\widetilde{\beta}_\vartheta$ shows a similar behaviour as $\widetilde{\alpha}_\vartheta$ in the previous section.

Lemma 3.1. *Assume (A1), (A2), (A3), $h_n \to 0$ and $\sup_y\{E(Z^2 \mid Y=y)\} < \infty$. Then, for n large enough,*

$$(i) \qquad |\beta_{\vartheta_0}| - |\beta_\vartheta| \geq \begin{cases} C_6|\vartheta_0 - \vartheta|^2/h_n - C_7 h_n|\vartheta_0 - \vartheta|, & \text{if } |\vartheta_0 - \vartheta| \leq h_n, \\ C_8 h_n, & \text{if } |\vartheta_0 - \vartheta| > h_n \end{cases},$$

$$(ii) \qquad \sqrt{var(\widetilde{\beta}_{\vartheta_0} - \widetilde{\beta}_\vartheta)} = O\left(n^{-1/2}|\vartheta_0 - \vartheta| h_n^{-\alpha-1/2}\right), \quad \text{if } |\vartheta_0 - \vartheta| \leq h_n$$

$$(iii) \qquad \sqrt{var(\widetilde{\beta}_\vartheta)} = O\left(n^{-1/2} h_n^{-\alpha+1/2}\right)$$

hold uniformly in $\vartheta \in [a,b]$ and $m \in \mathcal{F}$.

This lemma indicates that we can again estimate the location of the jump with a rate $\Delta_n = n^{-1/(\alpha+3/2)}$, which can actually be proved under the additional assumption

(A4) $\sup_y\{E(|Z|^M \mid Y=y)\} < \infty$ for M large enough.

Let, analogously to the definition of the change point estimator in the previous section,

$$\widehat{\vartheta}_0 \in \arg\max_{\vartheta\in[a,b]}\{|\widetilde{\beta}_\vartheta|\}. \tag{3.5}$$

This estimator converges again with the rate Δ_n to ϑ_0.

Theorem 3.1. *Assume (A1) through (A4) and $h_n \asymp n^{-1/(2\alpha+3)}$. Then*

$$\sup_{m\in\mathcal{F}}\left\{E_m\left(\Delta_n^{-1}(\widehat{\vartheta}_0 - \vartheta_0)\right)^2\right\} \leq C.$$

4. Lower bounds for change-point estimation

In "ordinary" change-point estimation it is known that the location of a jump can be estimated with the rate n^{-1} (or ϵ^2 in the closely related Gaussian white noise model); cf. Korostelev (1987). It can be shown that one can achieve this rate by a kernel-based method similar to that considered here. However, a special kernel has to be employed.

In view of these facts it is not obvious if the proposed method is optimal. To have an appropriate frame for such considerations, we consider minimax rates of convergence. Similarly to work of Fan (1991) and Fan and Truong (1993), who derived minimax rates for estimating the function itself in deconvolution and errors-in-variables regression, respectively, we obtain this minimax rate analyzing the Hellinger modulus of continuity at $n^{-1/2}$. As shown in Donoho and Liu (1987), this will immediately provide the desired lower bound for the rate of convergence.

4.1. A lower bound in the deconvolution problem. Our method consists in finding a sequence of two densities $f_{X,1,n}, f_{X,2,n} \in \mathcal{F}$ and some f_ε satisfying (A1), such that

$$H\left(f_{X,1,n} * f_\varepsilon, f_{X,2,n} * f_\varepsilon\right) = O(n^{-1/2}) \tag{4.1}$$

and

$$|\vartheta_{1,n} - \vartheta_{2,n}| \geq \Delta_n, \tag{4.2}$$

where $H(f,g) = \left(\int(\sqrt{f} - \sqrt{g})^2\right)^{1/2}$ is the Hellinger distance and $\vartheta_{i,n}$ are the change-points of $f_{X,i,n}$. From (4.1) and (4.2) we will infer that Δ_n is a lower bound for the minimax rate of convergence.

Now we consider two possible sets of conditions on $f_{X,i,n}$ and f_ε, which allow to derive (4.1) and (4.2). We try to assume as few as possible beyond (A1) for f_ε, and choose the $f_{X,i,n}$'s appropriately.

(A5) (i) f_ε satisfies (A1), and additionally $\int_{\mathbb{R}\setminus[c,d]} |f_\varepsilon^{(\beta)}(x)|\, dx \leq C$ for any integer $\beta \geq 2\alpha + 1$ and any compact interval $[c,d]$,
(ii) $f_{X,1,n} = f_{X,1} \in \mathcal{F}$ with $f_{X,1}(x) > 0$ for all x,
(iii) $f_{X,2,n} = f_{X,1,n} + b_n$, where the construction of b_n is exactly described in the proof of Lemma 4.1.

The essential features of b_n will be that $\int x^k b_n(x)\, dx = 0$ for $k = 0, 1, \ldots, \beta - 1$ and $\int |b_n(x)|\, dx = O(\Delta_n^{1/2})$.

To accommodate also the important problem of estimating the endpoint of the support of a density, we consider also the following restrictions on $f_{X,i,n}$ and f_ε.

(A5') (i) f_ε satisfies (A1), and additionally $\int_{\mathbb{R}\setminus[c,d]} |f_\varepsilon^{(\beta)}(x)|\, dx \leq C$ for any integer $\beta \geq 2\alpha + 1$ and any compact interval $[c,d]$, and $P(\varepsilon > d) > 0$,
(ii) $f_{X,1,n} = f_{X,1} \in \mathcal{F}$ with $f_{X,1}(x) > 0$ for all $x < \vartheta_{1,n}$,
(iii) $f_{X,2,n} = f_{X,1,n} + b_n$, where b_n is again defined in the proof of Lemma 4.1.

Now we obtain the following lemma.

Lemma 4.1. *Assume (A5) or (A5'). Then* (4.1) *and* (4.2) *are fulfilled.*

The final step towards the lower bound can be described as follows; cf. also Donoho and Liu (1987). Define for two densities p and q the Hellinger affinity $\rho(p,q) = \int \sqrt{p}\sqrt{q}$. Further, let $p^{[n]}$ denote the n-dimensional product measure of p. Then

$$\begin{aligned}
&\rho\left((f_{X,1,n} * f_\varepsilon)^{[n]}, (f_{X,2,n} * f_\varepsilon)^{[n]}\right) \\
&= \rho^n\left(f_{X,1,n} * f_\varepsilon, f_{X,2,n} * f_\varepsilon\right) \\
&= \left(1 - \frac{1}{2} H^2(f_{X,1,n} * f_\varepsilon, f_{X,2,n} * f_\varepsilon)\right)^n \\
&\geq (1 - C/n)^n \geq C > 0.
\end{aligned}$$

On the other hand, we have for any estimator $\widehat{\vartheta}_0$ that

$$\begin{aligned}
&\rho\left((f_{X,1,n} * f_\varepsilon)^{[n]}, (f_{X,2,n} * f_\varepsilon)^{[n]}\right) \\
&\leq \int \frac{|\widehat{\vartheta}_0(y) - \vartheta_{1,n}|}{\Delta_n}\sqrt{(f_{X,1,n} * f_\varepsilon)^{[n]}(y)}\sqrt{(f_{X,2,n} * f_\varepsilon)^{[n]}(y)}\, dy \\
&\quad + \int \frac{|\widehat{\vartheta}_0(y) - \vartheta_{2,n}|}{\Delta_n}\sqrt{(f_{X,1,n} * f_\varepsilon)^{[n]}(y)}\sqrt{(f_{X,2,n} * f_\varepsilon)^{[n]}(y)}\, dy \\
&\leq \sqrt{E_{f_{X,1,n}}\left(\Delta_n^{-1}(\widehat{\vartheta}_0 - \vartheta_{1,n})\right)^2} + \sqrt{E_{f_{X,2,n}}\left(\Delta_n^{-1}(\widehat{\vartheta}_0 - \vartheta_{2,n})\right)^2},
\end{aligned}$$

which immediately leads to the following theorem.

Theorem 4.1. *Assume (A5) or (A5'). Then*

$$\inf_{\widehat{\vartheta}_0} \max_{f_X \in \{f_{X,1,n}, f_{X,2,n}\}} \left\{E_{f_X}\left(\Delta_n^{-1}(\widehat{\vartheta}_0 - \vartheta_0)\right)^2\right\} \geq C > 0.$$

4.2. A lower bound for errors-in-variables regression. We will obtain the desired lower bound for the accuracy in estimating ϑ_0 again analyzing the Hellinger modulus of continuity of the problem. In the construction of the two densities $f_{(X,Z),1,n}$ and $f_{(X,Z),2,n}$ we adapt a nice idea of Fan and Truong (1993), which allows us to use the result about the Hellinger distance of the two densities $f_{X,1,n} * f_\varepsilon$ and $f_{X,2,n} * f_\varepsilon$ from the previous section.
To facilitate proofs, we choose f_ε and $f_{X,i,n}$ according to (A5). Further, we choose densities f_0, g_0 and a function h_0 with the following properties:

(A6) (i) $f_0(x) \geq C f_{X,1}(x) \quad \forall x$, $f_0(x) \equiv \bar{f}_0 > 0$ for $x \in [\vartheta_1 - \delta, \vartheta_1 + \delta]$ and any $\delta > 0$,
(ii) $\int z g_0(z)\, dz = 0$,
(iii) $\int h_0(z)\, dz = 0, \quad \int z h_0(z)\, dz = 1$,
(iv) $g_0(z) \geq 2\sup_x\{f_{X,1}(x)\}|h_0(z)| \quad \forall z$.

Now we define our two densities $f_{(X,Z),1,n}$ and $f_{(X,Z),2,n}$ as

$$f_{(X,Z),i,n}(x,z) = f_0(x)\left[g_0(z) + f_{X,i,n}(x)h_0(z)\right]. \tag{4.3}$$

Because of (A5)(ii), the marginal density of X turns out to be equal to f_0, which provides that

$$E_{f_{(X,Z),i,n}}(Z \mid X = x) = \int z[g_0(z) + f_{X,i,n}(x)h_0(z)]\, dz = f_{X,i,n}(x). \tag{4.4}$$

Hence, the regression function $m(x)$ has exactly the same change-points as the densities $f_{X,i,n}$ from the previous section. Finally, we can estimate the Hellinger distance between $f_{(Y,Z),1,n}$ and $f_{(Y,Z),2,n}$, where

$$f_{(Y,Z),i,n}(y,z) = \int f_{(X,Z),i,n}(y-x,z) f_\varepsilon(x)\, dx,$$

by

$$
\begin{aligned}
& H^2\left(f_{(Y,Z),1,n}, f_{(Y,Z),2,n}\right) \\
\leq\ & \int\int_{\vartheta_1+c-\delta}^{\vartheta_1+d+\delta} \frac{\left(f_{(Y,Z),1,n}(y,z) \,-\, f_{(Y,Z),2,n}(y,z)\right)^2}{f_{(Y,Z),1,n}(y,z)}\, dy\, dz \\
& + \int\int_{\mathbb{R}\setminus[\vartheta_1+c-\delta,\vartheta_1+d+\delta]} \left| f_{(Y,Z),1,n}(y,z) \,-\, f_{(Y,Z),2,n}(y,z)\right|\, dy\, dz \\
=\ & \int\int_{\vartheta_1+c-\delta}^{\vartheta_1+d+\delta} \frac{\left(h_0(z)\int f_0(y-x)b_n(y-x)f_\varepsilon(x)\, dx\right)^2}{\int f_{(X,Z),1,n}(y-x,z)f_\varepsilon(x)\, dx}\, dy\, dz \\
& + \int |h_0(z)|\, dz \int_{\mathbb{R}\setminus[\vartheta_1+c-\delta,\vartheta_1+d+\delta]} \left| \int f_0(y-x)b_n(y-x)f_\varepsilon(x)\, dx\right|\, dy \\
\leq\ & \frac{\bar{f_0}^2}{C} \int \frac{2(h_0(z))^2}{g_0(z)}\, dz \int \frac{\left(\int b_n(y-x)f_\varepsilon(x)\, dx\right)^2}{\int f_{X,1}(y-x)f_\varepsilon(x)\, dx}\, dy \\
& + O\left(\int_{\mathbb{R}\setminus[\vartheta_1+c-\delta,\vartheta_1+d+\delta]} \left| \int f_0(y-x)b_n(y-x)f_\varepsilon(x)\, dx\right|\, dy\right). \qquad (4.5)
\end{aligned}
$$

Here the essence of the idea of Fan and Truong becomes again apparent: we could separate the part involving integration over z from that which depends on y. The integrals over y on the right-hand side of (4.5) are just the same as I_1 and I_2 from the proof of Lemma 4.1.
Now we can derive, in analogy to Lemma 4.1, the following assertion.

Lemma 4.2. *Assume (A5) and (A6). Then*

$$H\left(f_{(Y,Z),1,n}, f_{(Y,Z),2,n}\right) = O(n^{-1/2}).$$

Hence, we have again the fact that the two experiments according to $f_{(Y,Z),1,n}$ and $f_{(Y,Z),2,n}$ are statistically not distinguishable. Arguing as above we get the desired lower bound for the accuracy in estimating ϑ_0.

Theorem 4.2. *Assume (A5) and (A6). Then*

$$\inf_{\widehat{\vartheta}_0} \max_{f_{(Y,Z)}\in\{f_{(Y,Z),1,n}, f_{(Y,Z),2,n}\}} \left\{ E_{f_{(Y,Z)}}\left(\Delta_n^{-1}(\widehat{\vartheta}_0 \,-\, \vartheta_0)\right)^2\right\} \geq C > 0.$$

5. Proofs

Lemma 5.1. *Assume (A2). Then*

(i) $\quad \|\gamma_\vartheta^{(k)}\|_2 = O(h_n^{-\alpha-k+1/2})$,
(ii) $\quad \|\gamma_\vartheta^{(k)}\|_\infty = O(h_n^{-\alpha-k})$

for $k = 0, 1, 2$.

Proof. We have that

$$\begin{aligned}
\|\gamma_\vartheta^{(k)}\|_2^2 &= (2\pi)^{-1}\||\omega|^k\widehat{\gamma_\vartheta}(\omega)\|_2^2 \\
&= O\left(\||\omega|^k\widehat{\varphi_\vartheta}(\omega)(1+|\omega|)^\alpha\|_2^2\right) \\
&= O\left(\|h_n\widehat{\varphi}(h_n\omega)(1+|\omega|)^\alpha|\omega|^k\|_2^2\right) \\
&= O\left(h_n\|\widehat{\varphi}(\omega)(1+|\omega|/h_n)^\alpha(|\omega|/h_n)^k\|_2^2\right) \\
&= O\left(h_n^{-2\alpha-2k+1}\right),
\end{aligned}$$

and, analogously,

$$\begin{aligned}
\|\gamma_\vartheta^{(k)}\|_\infty &\le (2\pi)^{-1}\||\omega|^k\widehat{\gamma_\vartheta}(\omega)\|_1 \\
&= O\left(\|h_n\widehat{\varphi}(h_n\omega)(1+|\omega|)^\alpha|\omega|^k\|_1\right) \\
&= O\left(h_n^{-\alpha-k}\|\widehat{\varphi}(\omega)(1+|\omega|)^\alpha(|\omega|)^k\|_1\right) \\
&= O\left(h_n^{-\alpha-k}\right).
\end{aligned}$$

□

Proof of Lemma 2.1.

(i) Let, w.l.o.g., $f_X(\vartheta_0-) = f_X(\vartheta_0+) + \Delta$ for some $\Delta > 0$. Note that f_X can be written as

$$f_X(x) = f(x) + h(x), \tag{5.1}$$

where f and h are uniformly Lipschitz on $\mathbb{R}$ and $\mathbb{R} \setminus \{\vartheta_0\}$, respectively. Moreover, we choose h such that $h(\vartheta_0-) = \Delta/2$, $h(\vartheta_0+) = -\Delta/2$, $h(x) \ge 0$, if $x < 0$, and $h(x) \le 0$, if $x > 0$ are satisfied.
Since φ_ϑ satisfies $\int \varphi_\vartheta(x)\,dx = 0$ and has a length of support of $O(h_n)$, we get

$$(\varphi_{\vartheta_0} - \varphi_\vartheta, f) = O\left(h_n \int |\varphi_{\vartheta_0}(x) - \varphi_\vartheta(x)|\,dx\right) = O\left((h_n|\vartheta_0 - \vartheta|) \wedge h_n^2\right). \tag{5.2}$$

Here the last equality holds, because φ has bounded total variation.
For $\vartheta_0 \in supp(\varphi_\vartheta)$ we obtain that

$$\begin{aligned}
(\varphi_{\vartheta_0} - \varphi_\vartheta, h) &= \Delta/2\int_{-\infty}^{\vartheta_0}(\varphi_{\vartheta_0}(x) - \varphi_\vartheta(x))\,dx - \Delta/2\int_{\vartheta_0}^{\infty}(\varphi_{\vartheta_0}(x) - \varphi_\vartheta(x))\,dx \\
&\quad + O\left(\int |x-\vartheta_0||\varphi_{\vartheta_0}(x) - \varphi_\vartheta(x)|\,dx\right), \\
&\ge C_6|\vartheta_0 - \vartheta|^2/h_n - C_7 h_n|\vartheta_0 - \vartheta|,
\end{aligned} \tag{5.3}$$

whereas we get for $\vartheta_0 \notin supp(\varphi_\vartheta)$ that

$$(\varphi_{\vartheta_0} - \varphi_\vartheta, h) \ge C_6 h_n - O\left(\int \varphi_\vartheta(x)h(x)\,dx\right) \ge C_6 h_n - C_7 h_n^2. \tag{5.4}$$

(i) follows from (5.2), (5.3) and (5.4).

(ii) Since $f_X * f_\varepsilon$ is bounded, we obtain that

$$var(\tilde{\alpha}_{\vartheta_0} - \tilde{\alpha}_\vartheta) \leq n^{-1} E\left(\gamma_{\vartheta_0}(Y) - \gamma_\vartheta(Y)\right)^2 = O\left(n^{-1} \int (\gamma_{\vartheta_0}(x) - \gamma_\vartheta(x))^2 \, dx\right).$$

Now we have by Lemma 5.1 that

$$\|\gamma_{\vartheta_0} - \gamma_\vartheta\|_2^2 = \int \left(\int_\vartheta^{\vartheta_0} \gamma'_\vartheta(x+z)\, dz \right)^2 dx \leq |\vartheta_0 - \vartheta|^2 \|\gamma'_\vartheta\|_2^2 = O(|\vartheta_0 - \vartheta|^2 h_n^{-2\alpha-1}),$$

which implies (ii).
(iii) follows from

$$var(\tilde{\alpha}_\vartheta) = O(n^{-1}\|\gamma_\vartheta\|_2^2) = O(n^{-1} h_n^{-2\alpha+1}).$$

□

Proof of Theorem 2.1. The proof of this theorem is mainly based on a repeated application of Bernstein's inequality, which we quote for reader's convenience from Shorack and Wellner (1986, p. 855):
Let $Z_1, \ldots, Z_n$ be i.i.d. random variables with $EZ_1 = 0$ and $|Z_1| \leq K$ almost surely. Then, for $\bar{Z} = n^{-1} \sum Z_i$,

$$\begin{aligned} P(\bar{Z} > c) &\leq \exp\left(-\frac{nc^2/2}{var(Z_1) + (Kc)/3}\right) \\ &\leq \exp\left(-\frac{c^2}{4\, var(\bar{Z})}\right) + \exp\left(-\frac{3nc}{4K}\right) \qquad (5.5) \end{aligned}$$

holds for arbitrary $c > 0$.
(i) Let $\xi_\vartheta = \tilde{\alpha}_\vartheta - \alpha_\vartheta$.
First we estimate the probability of the event $\tilde{\Omega} = \{\omega \mid \ |\xi_\vartheta| > |\alpha_{\vartheta_0}|/3$ for any $\vartheta \in [a,b]\}$. For that we approximate ξ_ϑ on the grid

$$\Gamma_n = \{kd_n \mid \ k \in \mathbb{Z}\} \cap [a,b],$$

where $d_n = o(h_n^{\alpha+2})$, $d_n^{-1} = O(n^\delta)$ for any $\delta < \infty$.
Since $|\alpha_{\vartheta_0}| \asymp h_n$, we obtain by Lemma 2.1, Lemma 5.1 and (5.5) for fixed ϑ that

$$\begin{aligned} P\left(|\xi_\vartheta| > |\alpha_{\vartheta_0}|/6\right) &\leq 2\exp\left(-\frac{|\alpha_{\vartheta_0}|^2/36}{4\, var(\xi_\vartheta)}\right) + 2\exp\left(-\frac{|\alpha_{\vartheta_0}| n}{8\|\gamma_\vartheta\|_\infty}\right) \\ &\leq 2\exp\left(-C\frac{h_n^2}{n^{-1}h_n^{-2\alpha+1}}\right) + 2\exp\left(-C\frac{h_n n}{h_n^{-\alpha}}\right) \\ &= O\left(\exp(-Cn^{1/(\alpha+3/2)}) + \exp(-Cn^{(\alpha+2)/(2\alpha+3)})\right). \qquad (5.6) \end{aligned}$$

Further, let $\vartheta^* = \vartheta^*(\vartheta) \in \Gamma_n$ be that element of Γ_n, which is closest to ϑ. Then we have with probability one that

$$|\xi_\vartheta - \xi_{\vartheta^*}| \leq 2\|\gamma_\vartheta - \gamma_{\vartheta^*}\|_\infty \leq 2d_n \|\gamma'_{\vartheta^*}\|_\infty = o(h_n) \leq |\alpha_{\vartheta_0}|/6 \qquad (5.7)$$

for all $\vartheta \in [a,b]$, if n is large enough. This yields

$$\begin{aligned} P(\tilde{\Omega}) &\leq P\left(|\xi_\vartheta - \xi_{\vartheta^*}| > |\alpha_{\vartheta_0}|/6 \quad \text{for any} \quad \vartheta \in [a,b]\right) + \sum_{\vartheta^* \in \Gamma_n} P\left(|\xi_{\vartheta^*}| > |\alpha_{\vartheta_0}|/6\right) \\ &= O\left(n^\delta \exp(-Cn^{1/(\alpha+3/2)})\right). \end{aligned} \tag{5.8}$$

(ii) It is obvious that $\omega \notin \tilde{\Omega}$ implies that

$$\hat{\vartheta}_0 \in \Theta = \{\vartheta \in [a,b] \mid \quad |\alpha_\vartheta| \geq |\alpha_{\vartheta_0}|/3\}.$$

We decompose Θ into the subsets

$$I_m = \{\vartheta \in \Theta \mid \quad \vartheta_m \leq \vartheta \leq \vartheta_{m+1}\}, \quad m = -M, \dots, 0, 1, \dots, M,$$

where $\vartheta_m = \vartheta_0 + m\Delta_n$ and M is the smallest possible integer such that Θ is covered by $\bigcup_{|m|\leq M} I_m$.
Provided n is large enough, we have by Lemma 2.1 that

$$\Theta \subseteq [\vartheta_0 - h_n, \vartheta_0 + h_n].$$

Moreover, since the "essential part" of α_ϑ, which is equal to

$$\Delta/2 \int_{-\infty}^{\vartheta_0} \varphi_\vartheta(x)\,dx \;-\; \Delta/2 \int_{\vartheta_0}^{\infty} \varphi_\vartheta(x)\,dx$$

(see the proof of Lemma 2.1 for details), is nonnegative for all ϑ, we get immediately that

$$|\alpha_\vartheta| \;=\; O(h_n^2), \quad \text{if} \quad sgn(\alpha_\vartheta) \neq sgn(\alpha_{\vartheta_0}).$$

Let, w.l.o.g., $\alpha_{\vartheta_0} > 0$. Then, again for n large enough,

$$\alpha_\vartheta > 0 \quad \text{for all} \quad \vartheta \in \Theta. \tag{5.9}$$

Hence, we obtain that

$$\Omega_m \;=\; \{\hat{\vartheta}_0 \in I_m\} \subseteq \{\Omega \mid \quad \xi_\vartheta - \xi_{\vartheta_0} \geq c_m \quad \text{for any} \quad \vartheta \in I_m\}, \tag{5.10}$$

where $c_m = \alpha_{\vartheta_0} \;-\; \sup_{\vartheta \in I_m}\{\alpha_\vartheta\}$. From Lemma 2.1 we have that

$$c_m \;\geq\; C(m\Delta_n)^2/h_n \quad \text{for} \quad m_0 \leq |m| \leq M, \tag{5.11}$$

and appropriately fixed m_0. Let $\vartheta_{m,k} = \vartheta_m + (k/L_n)\Delta_n$, $k = 0, \dots, L_n - 1$, where L_n will be chosen below. For $\vartheta \in [\vartheta_{m,k}, \vartheta_{m,k+1}) \subseteq [\vartheta_m, \vartheta_{m,1})$ we consider the decomposition

$$\xi_\vartheta - \xi_{\vartheta_0} \;=\; (\xi_{\vartheta_m} - \xi_{\vartheta_0}) \;+\; (\vartheta_{m,k} - \vartheta_m)\xi'_{\vartheta_m} \;+\; \int_{\vartheta_m}^{\vartheta_{m,k}} (\xi'_z - \xi'_{\vartheta_m})\,dz \;+\; (\xi_\vartheta - \xi_{\vartheta_{m,k}}). \tag{5.12}$$

By Bonferroni's inequality we obtain that

$$\begin{aligned}
P(\Omega_m) \;&\leq\; P\left(\xi_{\vartheta_m} - \xi_{\vartheta_0} \geq c_m/4\right) \\
&\quad + P\left(\Delta_n \xi'_{\vartheta_m} \geq c_m/4\right) \\
&\quad + P\left(\xi_{\vartheta_{m,k}} - \xi_{\vartheta_m} - (\vartheta_{m,k} - \vartheta_m)\xi'_{\vartheta_m} \geq c_m/4 \quad \text{for any} \;\; k \in 0, \ldots, L_n - 1\right) \\
&\quad + P\left(\xi_\vartheta - \xi_{\vartheta_{m,k}} \geq c_m/4 \quad \text{for any} \;\; k \in 0, \ldots, L_n - 1, \;\; \vartheta \in [\vartheta_{m,k}, \vartheta_{m,k+1})\right) \\
&= \; p_{m1} \,+\, \ldots \,+\, p_{m4}.
\end{aligned} \tag{5.13}$$

It will turn out that this somewhat involved decomposition (5.12) is really necessary for proving that $P(\Omega_m)$ decreases fast enough as $|m|$ grows. To get a valid result for a continuum of points, we have to approximate $\xi_\vartheta - \xi_{\vartheta_0}$ on a sufficiently fine grid, with spacings that are of smaller order of magnitude than Δ_n. Hence, even on a fixed interval $I_m = [\vartheta_m, \vartheta_{m+1})$ we have to consider an increasing number of grid points. The stochastic fluctuations of the first two terms on the right-hand side of (5.13) will be of about the same order of magnitude as c_m for small $|m|$; hence we consider only a single term of these kinds for every m. The third term is of smaller order of magnitude than c_m; therefore we can include an increasing number of them (namely L_n for every m). Finally, provided we choose L_n large enough, we find a non-stochastic upper estimate for the fourth term, which is of smaller order than c_m.
Now we turn to the estimation of p_{m1} through p_{m4}. Using

$$var(\xi_{\vartheta_m} - \xi_{\vartheta_0}) \;=\; O\left(n^{-1}|\vartheta_m - \vartheta_0|^2 h_n^{-2\alpha-1}\right) \;=\; O((m\Delta_n)^2 h_n^2),$$

$$\|\gamma_{\vartheta_m} - \gamma_{\vartheta_0}\|_\infty \;=\; O\left(|\vartheta_m - \vartheta_0| h_n^{-\alpha-1}\right) \;=\; O(m\Delta_n h_n^{-\alpha-1}),$$

and (5.11) we obtain due to (5.5) that

$$\begin{aligned}
p_{m1} \;&\leq\; \exp\left(-\frac{(c_m/4)^2}{4\, var(\xi_{\vartheta_m} - \xi_{\vartheta_0})}\right) \,+\, \exp\left(-\frac{3nc_m/4}{4\|\gamma_{\vartheta_m} - \gamma_{\vartheta_0}\|_\infty}\right) \\
&\leq\; \exp(-Cm^2) \,+\, \exp(-Cmn^{(\alpha+1)/(2\alpha+3)}).
\end{aligned} \tag{5.14}$$

By

$$var(\Delta_n \xi'_{\vartheta_m}) \;=\; O\left(n^{-1}\Delta_n^2 \|\gamma'_{\vartheta_m}\|_2^2\right) \;=\; O(\Delta_n^2 h_n^2)$$

and

$$\Delta_n \|\gamma'_{\vartheta_m}\|_\infty \;=\; O\left(\Delta_n h_n^{-\alpha-1}\right) \;=\; O(\Delta_n h_n^{-\alpha-1}),$$

we get analogously to (5.14) that

$$p_{m2} \,\leq\, \exp(-Cm^4) \,+\, \exp(-Cm^2 n^{(\alpha+1)/(2\alpha+3)}). \tag{5.15}$$

Further, we obtain from

$$var\Big(\int_{\vartheta_m}^{\vartheta_{m,k}} (\xi'_z - \xi'_{\vartheta_m})\, dz\Big) \;=\; O\left(n^{-1}\left\|\int_{\vartheta_m}^{\vartheta_{m,k}} (\gamma'_z - \gamma'_{\vartheta_m})\, dz\right\|_2^2\right) \;=\; O\left(n^{-1}\Delta_n^4 \|\gamma''_\vartheta\|_2^2\right) \;=\; O\left(\Delta_n^4\right)$$

and

$$\left\|\int_{\vartheta_m}^{\vartheta_{m,k}} (\gamma'_z - \gamma'_{\vartheta_m})\, dz\right\|_\infty \;=\; O\left(\Delta_n^2 \|\gamma''_\vartheta\|_\infty\right) \;=\; O\left(\Delta_n^2 h_n^{-\alpha-2}\right)$$

that

$$\begin{aligned} p_{m3} &\leq L_n \exp\left(-C\frac{(m\Delta_n)^4/h_n^2}{\Delta_n^4}\right) + L_n \exp\left(-C\frac{n(m\Delta)^2/h_n}{\Delta^2 h_n^{-\alpha-2}}\right) \\ &\leq L_n \exp(-Cm^4 n^{2/(2\alpha+3)}) + L_n \exp(-Cm^2 n^{(\alpha+2)/(2\alpha+3)}). \end{aligned} \tag{5.16}$$

Finally, we get for $|\vartheta - \vartheta_{m,k}| = O(L_n^{-1})$ that $\xi_\vartheta - \xi_{\vartheta_{m,k}} = O(L_n^{-1}\|\gamma'_\vartheta\|_\infty) = o(c_m)$, if $L_n^{-1} = o(h_n^{\alpha+4}) = o(n^{-(\alpha+4)/(2\alpha+3)})$. Provided we choose L_n in such a way, we obtain

$$p_{m4} = 0. \tag{5.17}$$

Now we conclude from (5.8) and (5.14) through (5.17) that

$$\begin{aligned} E\left(\Delta_n^{-1}(\widehat{\vartheta}_0 - \vartheta_0)\right)^2 &\leq O(\Delta_n^{-2})P(\tilde{\Omega}) \\ &\quad + (|m_0|+1)^2 \\ &\quad + \sum_{|m|=m_0}^{M} (|m|+1)^2 P(\Omega_m) \\ &= O(1). \end{aligned}$$

□

Proof of Theorem 3.1. This theorem can be proved in complete analogy to Theorem 2.1. The only modification will concern the exponential decay of the probability bounds in (5.8) and (5.14) through (5.16). Since we assume only that a certain finite number of moments of Z are finite, we will get an additional term of order $O(n^{-\lambda})$ in these bounds, where λ can be chosen arbitrarily large in dependence on a sufficiently large choice of M in (A4). Instead of Bernstein's inequality we can then apply results of Nagaev (1965). □

Proof of Lemma 4.1.
(i) Construction of b_n
Let p_i be Lipschitz continuous functions supported on $[-1,0]$ such that

$$\int x^j p_i(x)\,dx = -\delta_{ij}/(i+1) \quad \text{for } j = 0,1,\dots,\beta-1.$$

Now we define

$$b_n(x) = \Delta \sum_{i=0}^{\beta-1} \Delta_n^{(i+1)/2} p_i\left(\frac{x-\vartheta_1}{\sqrt{\Delta_n}}\right) + \Delta I(\vartheta_1 \leq x \leq \vartheta_1 + \Delta_n). \tag{5.18}$$

It is easy to see that b_n is supported on $[\vartheta_1 - \sqrt{\Delta_n}, \vartheta_1 + \Delta_n]$, satisfies $\int |b_n(x)|\,dx = O(\Delta_n)$, and $\int x^j b_n(x)\,dx = 0$ for $j = 0,1,\dots,\beta-1$.
(ii) Proof of (4.1) under (A5)

We have

$$\begin{aligned} & H^2\left(f_{X,1} * f_\varepsilon, f_{X,2,n} * f_\varepsilon\right) \\ \leq\ & \int_{\vartheta_1+c-\delta}^{\vartheta_1+d+\delta} \frac{((b_n * f_\varepsilon)(x))^2}{(f_{X,1} * f_\varepsilon)(x)}\, dx \ +\ \int_{\mathbb{R}\setminus[\vartheta_1+c-\delta,\vartheta_1+d+\delta]} |(b_n * f_\varepsilon)(x)|\, dx \\ =\ & I_1 \ +\ I_2. \end{aligned} \tag{5.19}$$

Since $f_{X,1}(x) > 0 \quad \forall x$ we have that $(f_{X,1} * f_\varepsilon)(x) > 0 \quad \forall x$, which means that $f_{X,1} * f_\varepsilon$ is bounded away from zero on the compact interval $[\vartheta_1 + c - \delta, \vartheta_1 + d + \delta]$. Hence, we obtain that

$$\begin{aligned} I_1 \ &\leq\ C \int ((b_n * f_\varepsilon)(x))^2\, dx \\ &=\ C \int |\widehat{b_n}(\omega)\widehat{f}_\varepsilon(\omega)|^2\, d\omega \\ &\leq\ C\Delta_n \int \left| \widehat{b_n(./\sqrt{\Delta_n})}(\sqrt{\Delta_n}\omega)|\omega|^{-\alpha} \right|^2 d\omega \\ &=\ C\Delta_n^{\alpha+1/2} \int \left| \widehat{b_n(./\sqrt{\Delta_n})}(\omega)|\omega|^{-\alpha} \right|^2 d\omega \\ &\leq\ C\Delta_n^{\alpha+1/2} \left[\int |b_n^{(-1)}(x/\sqrt{\Delta_n})|^2\, dx \ +\ \int |b_n^{(-(\beta-1))}(x/\sqrt{\Delta_n})|^2\, dx \right] \\ &=\ O\left(\Delta_n^{\alpha+3/2}\right) \ =\ O(n^{-1}). \end{aligned} \tag{5.20}$$

From a Taylor expansion with integral remainder we get that

$$\begin{aligned} (b_n * f_\varepsilon)(x) \ &=\ \int b_n(y) f_\varepsilon(x-y)\, dy \\ &=\ O\left(\Delta_n^{(\beta-1)/2} \int |b_n(y)|\, dy \int_{y \in supp(b_n)} |f_\varepsilon^{(\beta)}(x-y)|\, dy \right) \end{aligned}$$

holds for $x \in \mathbb{R} \setminus [\vartheta_1 + c - \delta, \vartheta_1 + d + \delta]$, which implies that

$$\begin{aligned} I_2 \ &\leq\ C\Delta_n^{(\beta+1)/2} \int_{\mathbb{R}\setminus[\vartheta_1+c-\delta,\vartheta_1+d+\delta]} \int_{y\in supp(b_n)} |f_\varepsilon^{(\beta)}(x-y)|\, dy\, dx \\ &=\ O\left(\Delta_n^{(\beta+2)/2}\right) \ =\ O(n^{-1}). \end{aligned} \tag{5.21}$$

This completes the proof of (4.1). The proof under (A5') is similar, since $(f_{X,1} * f_\varepsilon)(x) \geq C > 0$ can also be shown for $x \in [\vartheta_1 + c - \delta, \vartheta_1 + d + \delta]$ and sufficiently small $\delta > 0$.
(4.2) follows directly from the construction of $f_{X,2,n}$. □

Acknowledgment. I thank Jianqing Fan for helpful discussions.

References

1. Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184–1186.
2. Donoho, D. L. (1992). Nonlinear solution of linear inverse problems by Wavelet-Vaguelette Decomposition. Technical Report, Department of Statistics, Stanford University.
3. Donoho, D. L. and Liu, R. C. (1987). Geometrizing rates of convergence, I., Technical Report 137a, Department of Statistics, University of California, Berkeley.
4. Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.
5. Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21**, 1900–1925.
6. Hall, P. (1990). Optimal convergence rates in signal recovery. *Ann. Statist.* **18**, 887–900.
7. Hall, P. and Titterington, D. M. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics* **34**, 429–440.
8. Korostelev, A. P. (1987). On minimax estimation of a discontinuous signal. *Theory Probab. Appl.* **32**, 727–730.
9. Müller, H. G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20**, 737–761.
10. Nagaev, S. V. (1965). Some limit theorems for large deviations. *Theory Probab. Appl.* **10**, 214–235.
11. Shorack, G. R. and Wellner, J. A. (1986). Empirical Processes with Applications to Statistics. Wiley, New York.
12. Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82**, 385–397.
13. Wu, J. S. and Chu, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. *Ann. Statist.* **21**, 1545–1566.
14. Yin, Y. Q. (1988). Detection of the number, locations and magnitudes of jumps. *Comm. Statist. Stochastic Models* **4**, 445–455.
15. Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806–831.